BBA - Gene Regulatory Mechanisms 1864 (2021) 194749

Contents lists available at ScienceDirect

BBA - Gene Regulatory Mechanisms

journal homepage: www.elsevier.com/locate/bbagrm

Integration of transcription coregulator complexes with sequence-specific DNA-binding factor interactomes☆

Niels Velthuijs [a,1], Birgit Meldal [b,c,1], Quinte Geessinck [a], Pablo Porras [b,c], Yulia Medvedeva [d], Anatoliy Zubritskiy [d], Sandra Orchard [b], Colin Logie [a,*]

[a] Molecular Biology Department, Science Faculty, Radboud University, PO Box 9101, 6500HB Nijmegen, the Netherlands
[b] European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK
[c] Open Targets, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK
[d] Institute of Bioengineering, Research Center of Biotechnology, Russian Academy of Sciences, Moscow 119071, Russian Federation

ARTICLE INFO

Keywords:
Transcription regulation
dbTF
Nuclear receptor
TF coregulators
Chromatin remodelling
Complex Portal

ABSTRACT

The domain of transcription regulation has been notoriously difficult to annotate in the Gene Ontology, partly because of the intricacies of gene regulation which involve molecular interactions with DNA as well as amongst protein complexes. The molecular function 'transcription coregulator activity' is a part of the biological process 'regulation of transcription, DNA-templated' that occurs in the cellular component 'chromatin'. It can mechanistically link sequence-specific DNA-binding transcription factor (dbTF) regulatory DNA target sites to coactivator and corepressor target sites through the molecular function 'cis-regulatory region sequence-specific DNA binding'. Many questions arise about transcription coregulators (coTF). Here, we asked how many unannotated, putative coregulators can be identified in protein complexes? Therefore, we mined the CORUM and hu.MAP protein complex databases with known and strongly presumed human transcription coregulators. In addition, we trawled the BioGRID and IntAct molecular interaction databases for interactors of the known 1457 human dbTFs annotated by the GREEKC and GO consortia. This yielded 1093 putative transcription factor coregulator complex subunits, of which 954 interact directly with a dbTF. This substantially expands the set of coTFs that could be annotated to 'transcription coregulator activity' and sets the stage for renewed annotation and wet-lab research efforts. To this end, we devised a prioritisation score based on existing GO annotations of already curated transcription coregulators as well as interactome representation. Since all the proteins that we mined are parts of protein complexes, we propose to concomitantly engage in annotation of the putative transcription coregulator-containing complexes in the Complex Portal database.

1. Introduction

In its simplest form, eukaryotic transcription initiation only requires the general transcription machinery. This machinery consists of RNA polymerase II and the general transcription initiation factors (GTFs) TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH [1,2]. These bind promoter DNA sequences and promoter DNA-bound proteins and recruit the RNA polymerase to form the transcription preinitiation complex (PIC) [3]. After formation of the PIC, transcription can be initiated by activation of RNA polymerase. This is followed by promoter escape and transcription elongation [4]. All this requires a multitude of nucleosome remodelling factors, many of which have partially overlapping and redundant molecular functions [5]. An overview of the functional distinctions between proteins acting in transcription and transcription regulation is shown in Fig. 1.

To express all the proteins required for morphological and functional cellular differentiation of all of the human cell types, elaborate transcriptional regulation evolved, as demonstrated by the positive correlation between higher organismal complexity and transcription regulation intricacy [6]. Gene-specific transcription regulation requires

☆ This article is part of a Special Issue entitled: Curation of the Gene Regulatory Knowledge Commons edited by Dr. Colin Logie, Dr. Wyeth Wasserman and Dr. Julio Collado.

* Corresponding author.
E-mail addresses: bmeldal@ebi.ac.uk (B. Meldal), pporras@ebi.ac.uk (P. Porras), orchard@ebi.ac.uk (S. Orchard), c.logie@science.ru.nl (C. Logie).

[1] These authors contributed equally.

https://doi.org/10.1016/j.bbagrm.2021.194749
Received 10 February 2021; Received in revised form 13 August 2021; Accepted 19 August 2021
Available online 21 August 2021

the regulatory machinery to recognize, and act at, specific DNA sequences. Sequence-specific double stranded DNA-binding transcription factors (dbTFs) are proteins that bind specific DNA sequences in regulatory regions of genes such as promoters and enhancers to regulate gene transcription. The paradigm for the mechanism of action of transcription factors is to bind to a gene regulatory sequence and to consequently regulate gene transcription by direct binding or via interactions with coregulating proteins that stimulate or repress a step of transcription regulation [7,8]. This latter step may be a result of interference with the binding of an activating transcription factor by binding and thus competing for the same regulatory DNA sequence or directly interfering with the productive binding of a GTF at a promoter [9]. Thus, the modularity of transcription factors and regulatory elements gives rise to many different mechanisms of action, and it has rarely been possible to study concomitant versus consecutive activities that potentially result in a multitude of context- and cell type-specific transcription regulation programs, that are ultimately idiosyncratic [10].

The dbTFs often need assistance from additional proteins to regulate transcription. These are termed transcription coregulators (coTFs). One mechanistic function of coTFs is to bridge the DNA-bound transcription factor to the general transcription machinery. Mediator, a large complex comprising 30 subunits in humans, is a well-known coTF that bridges enhancer-bound transcription factors to the GTFs [11,12]. A second important mechanism of action through which coTFs regulate transcription is by recruiting other coTFs or coregulating protein complexes that modify chromatin structure [13,14]: chromosomal DNA is packaged into highly structured chromatin, with the DNA wrapped in nucleosomes which generally occludes the DNA for DNA-binding proteins [15]. The accessibility of the DNA for transcription factors, GTFs and other chromatin modifying complexes, and thereby the transcription of specific genes, can be regulated by placement of specific modifications such as mono-, di- and trimethylation or acetylation on different amino acids of the histone tails [16,17]. Additionally, movement of nucleosomes [18,19], exchange of histone variants [20], and methylation of DNA itself can regulate transcription [21,22].

The Gene Ontology (GO) provides the most comprehensive knowledgebase on gene functions [23,24]. Gene products are annotated both manually and computationally with terms describing their function. Every GO term is sharply defined and falls into one of the three GO aspects, namely, *molecular function* (MF), *biological process* (BP) or *cellular component* (CC). The *molecular function* annotations describe inherent activity of a gene product, such as enzymatic activities or binding activities. The *cellular component* describes the location where the gene product performs its molecular function(s). This can be a cellular compartment but also a specific protein complex. Finally, the *biological process* aspect describes the larger pathway in which the gene product acts. Currently, GO has associated almost 1500 human proteins with the dbTF (GO:0003700) and 500 with the coTF (GO:0003712) molecular function terms or one of their descendants. But GO continually adds annotations. Since the advent of full genome sequences, potential dbTFs have been found computationally by searching for proteins with DNA-binding domains [25–27, Lovering et al., BBA this issue]. Biochemical assays are then required to demonstrate DNA motif-specific binding in addition to consequent gene transcription regulation activity. Finding potential coTFs needs a more elaborate approach. Here we deploy a computational approach; we asked how many human coTFs reside in protein complexes and accessorily also physically contact a known dbTF. Our results should empower wet-lab approaches as well as curation of currently known protein complexes that harbour coTFs. The accessory links we find between coTFs and dbTFs that may punctually recruit, activate or repress coTF activity at target genes form a basis to link genomic transcription regulatory chromosomal DNA sequences to the transcription regulatory proteins that act on them, and thus enable mechanistic modelling of the signal transduction pathways that lead to the process of transcriptional gene regulation. Since there appear to be almost as many coTFs as dbTFs, a crucial step in such modelling will

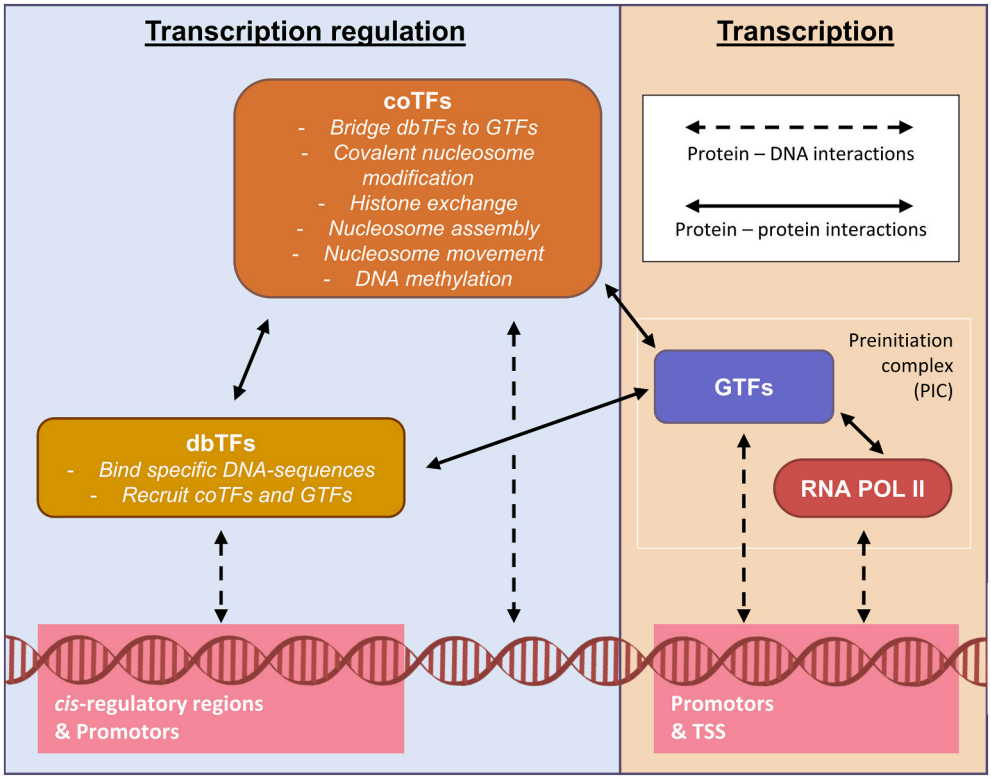


Fig. 1. Overview of the processes of transcription and its regulation. Firstly, there are the genomic address-recognising dbTFs that provide the system with gene-specificity by virtue of sequence-specific DNA binding. Secondly, there are the coTFs that modify chromatin structure or bridge the transcription factors to the general machinery. Finally, there are the general machinery proteins whose activities are required for transcription of all protein-coding genes by RNA polymerase II.

involve ascribing GO molecular functions to protein complexes, to enable subjugation of subunit functions [Cortés et al., BBA this issue]. We therefore propose to concomitantly engage in annotation of the putative transcription coregulator-containing complexes in the Complex Portal [28], since it already hosts many canonical transcription regulatory protein complexes and their annotations.

2. Material and methods

2.1. Code and data source versions

All code is accessible in GitHub repository https://github.com/nvelthuijs/cofactormining. Our algorithms were written using Python 3.6.5 through Jupyter Notebook. A flow chart of the cybernetic screens we performed is shown on Fig. 2. For all GO annotations and the Complex Portal, the most current version as available on 2020-11-01 was used. For all other sources the most current version as available on 2020-04-03 was used.

2.2. Sequential expansion of three bait lists

To mine the data sources for putative coTFs (preys), three successively expanded bait sets were compiled (Fig. 2A). Set 1 are the known coTFs, consisting of 419 human proteins associated with the GO:0003712 'transcription coregulator activity' annotation by the Gene Ontology Consortium. This included the descendant GO *molecular function* terms GO:0003713 'transcription coactivator activity' and GO:0003714 'transcription corepressor activity'. Twenty-one coTFs were excluded from bait set 1 because they are also annotated as dbTFs [Lovering et al., BBA this issue]. Those coTFs were included in bait set 3 (see below).

Set 2 encompasses bait set 1 plus four independently compiled lists of presumed coTFs. Firstly, 316 coTFs curated by the Nuclear Receptor Signaling Atlas (NURSA) [29,30]. Secondly, a positive control set (PCS), resulting from in-house curation of the deeply conserved eukaryotic general transcription factors reviewed by [31] as well as a set of transcription regulator complexes studied at the molecular biology department of the Radboud University in the course of the last two decades, comprising 391 proteins from 68 complexes. The 27 dbTFs in the PCS were left out. Third and fourth, 2 sets of manually identified chromatin remodelling enzymes from the UniProt knowledgebase: 134 enzymes known to covalently modify histones or DNA and a list of 22 SNF2-type ATPases that remodel nucleosomes. Due to the overlap amongst these in-house lists as well as with bait set 1, bait set 2 adds up to 893 proteins (see Fig. 3).

Bait set 3 counts 2341 proteins and consists of bait set 2 plus the 1457 human dbTFs annotated to GO:0003700 'DNA-binding transcription factor activity' and descendant terms, as recently defined by the GO consortium in collaboration with GREEKC members [Gaudet et al. - BBA this issue].

2.3. dbTF interacting protein resources

Two comprehensive resources of macromolecular complexes were mined for complexes that contain at least one of the 2341 proteins from the complete baits set: the Human Protein Complex Map (hu.MAP, http://proteincomplexes.org/) [32] and the Comprehensive resource of mammalian protein complexes (CORUM, mips.helmholtz-muenchen.

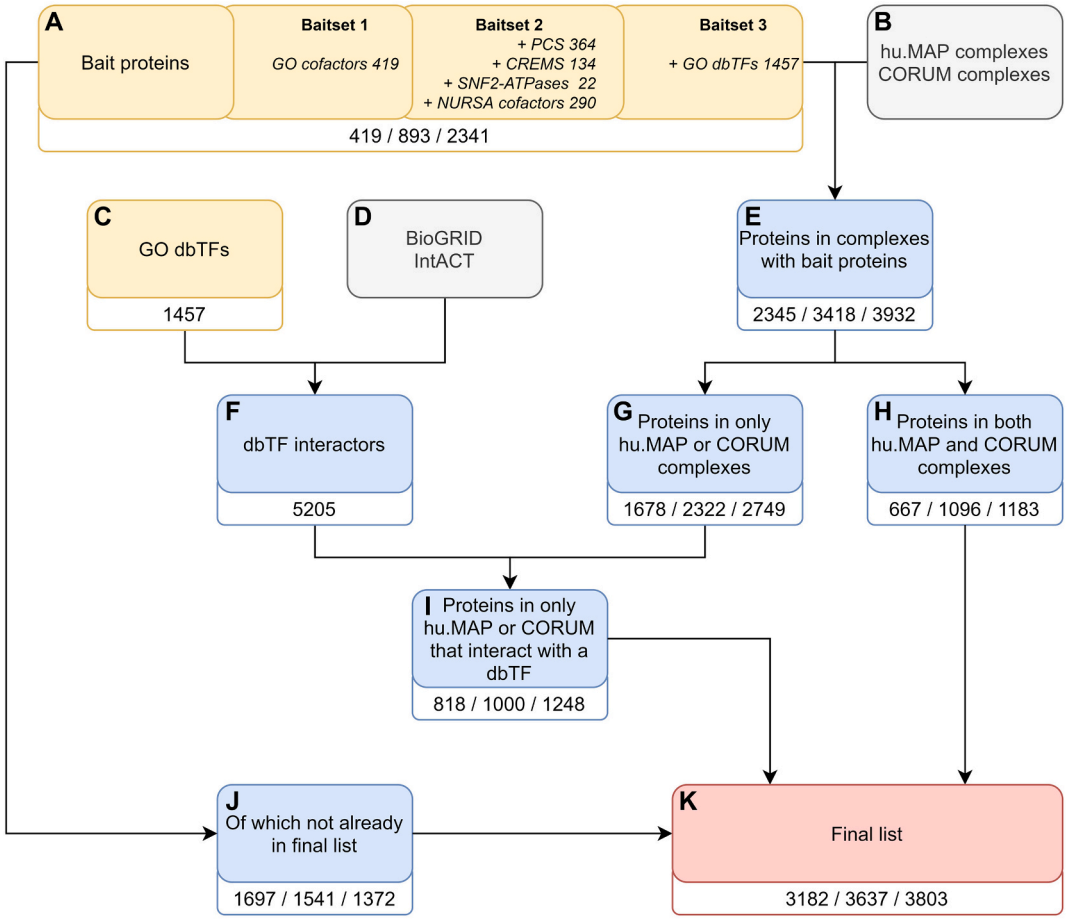


Fig. 2. Workflow for selection of putative coTFs. Numbers below each box show the number of proteins that resulted from that step for each of the three -nested- bait sets we used.

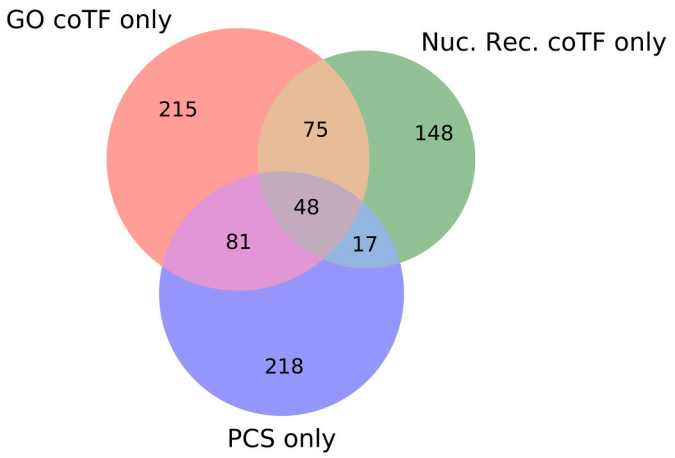


Fig. 3. Venn diagram for (i) the 419 known coTFs annotated by the Gene Ontology Consortium to GO:0003712 or a descendant MF term, (ii) 288 nuclear receptor coTFs from the NURSA database and (iii) 364 proteins residing in 68 'in house' positive control complexes (PCS). Note the substantial overlap between these known and presumed coTF lists that make-up bait set 2.

de/corum/) [33]. Both resources contain about 5000 complexes. For simplicity, we applied the matrix expansion model assuming that all subunits in a complex interact with each other simultaneously. Furthermore, two molecular interaction databases that supply molecular protein-protein interactions, BioGRID [34] (https://thebiogrid.org/) and IntAct [35] (www.ebi.ac.uk/intact), were mined for protein-protein interactions with the 1457 human dbTFs listed in the GO catalogue [Lovering et al., BBA this issue].

2.4. Mining the data sources

First, all proteins in complexes with bait proteins according to hu.MAP or CORUM were extracted from these sources (Fig. 2E). All proteins that were found in both sources were directly added to the final list of prey, (Fig. 2H, K). Second, all proteins in an interaction with a dbTF were extracted from BioGRID and IntAct (Fig. 2F). To retain an interaction detected in BioGRID we required that a protein-protein interaction be supported by two literature references. For IntAct we used a MI-SCORE threshold of 0.4, equivalent to evidence being recorded from two publications or from strong and validated evidence, using generally more than one experimental methodology, from one paper [36]. Thirdly, proteins that were found only in hu.MAP or CORUM (Fig. 2G) but for which there is additional protein-protein interaction evidence with a dbTF in either BioGRID or IntAct were added to the prey list (Fig. 2I). Fourth, any bait protein that did not make the final list through the above approaches was added to the prey list (Fig. 2J). The final list can be divided into a number of subsets: 'newly found coTFs' are all preys on the final list not already present as baits; 'putative coTFs' are all proteins on the final list not already annotated with dbTF or coTF GO terms (SupTab1). The latter set is larger than the newly found proteins, as it includes bait proteins from bait set 2 that are not (yet) annotated as a dbTF or coTF in the Gene Ontology.

2.5. GO score

To evaluate the current annotations of the putative coTFs, all GO annotations from the 440 human proteins that are annotated to GO:0003712 'transcription coregulator activity' were extracted. Only GO terms, as well as all their child terms, that indicate either involvement in the biochemistry of transcription or its regulation as we know it to take place in vivo were included in further analyses.

The final selection was a set of 478 GO terms, consisting of 89, 108 and 276 terms in the GO aspects *cellular component*, *molecular function* and *biological process*, respectively (SupTab2). A weight between 0.5 and 1 was calculated for each of these selected GO terms based on the fraction of proteins co-annotated to 'GO:0003712 transcription coregulator activity' and the selected GO term. A selected term that is associated with all known transcription coregulators that are associated with 'GO:0003712 transcription coregulator activity', or a descendant term, has a weight of 1, while a selected term that is not associated with any of the known transcription coregulators in GO has a weight of 0.5. This lower limit of 0.5 per term was set because some terms, for example the protein complex 'GO:0140091, mBAF complex', are very specific and therefore not used very often, resulting in a weight close to zero 0 if no lower limit were imposed, even though such terms strongly point to a transcription coregulator function, warranting a higher weight. A GO score was then calculated within each aspect (BP, MF and CC), by summing the weight of the five highest weighted terms, or all terms if a protein had less than five selected annotations in that GO aspect. If a protein did not have any of the selected terms annotated in an aspect, a penalty score of −1 was given for every not-selected annotation in that aspect, down to a maximum penalty of −5 per aspect. For each protein the penalties and scores from the three aspects were then added up, resulting in a final GO score with a theoretical range of −15 to +15. A score of −15 would mean that a protein has no selected transcription regulation-related terms associated with it in any of the GO-aspects CC, MF and BP, but that it does have at least five other annotations in each aspect. A score above 10 would mean that a protein has multiple highly weighted terms annotated in each of the three GO-aspects. The GO-scoring procedure is schematically rendered in Fig. 5.

2.6. Mining score

The mining score is a score between 0 and 5 that counts in how many of the starting data sources (bait proteins, hu.MAP, CORUM, IntAct and BioGRID) each protein was found as a potential transcription regulator. Because we wished to prioritize coTFs, the mining point for being a bait protein was only awarded for being a bait protein in bait sets 1 and 2. The mining score therefore reflects that a known, presumed or putative coTF protein resides in a complex or interacts with known coTFs or dbTFs.

2.7. Prioritisation score

To come to a final score, the GO score was first divided by 3 to make it fall within the range [−5, 5] to make it comparable to the mining score [0, 5]. Then the GO and mining scores were added to come to a final score with the theoretical range [−5, 10]. The complete scoring system is schematically outlined in Fig. 5.

2.8. Blacklist filter

Because some bait proteins have functions other than being coTFs the list of preys includes proteins from complexes that are known not to be coTFs, in particular the three largest protein complexes present inside the nucleus, the spliceosome, the pre-ribosome and proteasome. To allow for optional exclusion of these proteins during analysis, proteins annotated with GO terms *Cellular Component* ribosome (GO:0005840, GO:0022625, GO:0005762, GO:0022627, GO:0005763, GO:0015935, GO:0005761, GO:0015934, GO:0030687), spliceosome (GO:0071013, GO:0005681, GO:0071011, GO:0005689, GO:0071004) or proteasome (GO:0000502) as well as the biological process GO:0000398 (mRNA splicing, via spliceosome) were marked so as to permit blacklisting for prioritisation analyses (SupTab1).

3. A cybernetic screen for putative coTF complex subunits

The hu.MAP resource is the result of high-level computational integration of over 9000 mass-spectrometry experiments, generating over

4000 human protein complexes [32]. CORUM is a manually curated database that also contains more than 4000 mammalian protein complexes [33]. To discover putative coTF complex subunits, we mined the hu.MAP and CORUM human protein complex databases with three sets of 'bait' proteins, successively adding known and presumed coTFs and finally including all the known dbTFs as baits. In parallel we mined the molecular interaction databases BioGRID and IntAct for interaction partners of the known 1457 dbTFs (Fig. 2F). When a protein was discovered in both hu.MAP and CORUM, it was automatically included in the final list of putative coTFs. When it was only present in one of the two resources, we included it if it also interacted directly with one of the known dbTFs [Lovering et al. BBAGRM-D-20-00141 this issue] in the BioGRID or IntAct databases (Fig. 2H). This assumes that any protein that forms a stable complex with a known coTF, or that interacts directly with a dbTF, is likely to be a coTF. The results of the three successive screens performed with the successively expanded bait sets are quantified on Fig. 2 panels E–K.

3.1. Recovery of CORUM and hu.MAP coTF complexes

We set to produce a list of putative coTFs, under the assumption that all the subunits of a protein complex harbouring a known [23] or presumed [30,31] coTF are putative coTFs. Bait set 1 yielded 2345 proteins from either hu.MAP or CORUM, including the 419 known coTF bait proteins annotated as such in the Gene Ontology. Bait set 2 recovered an additional 1073 proteins. Bait set 3 included all the human dbTFs curated by GO and yielded only 166 additional proteins (Fig. 2E). However, when used by themselves, the 1457 GO dbTFs recovered 1942 proteins from hu.MAP and/or CORUM (not shown). This suggests that bait set 3 is close to saturation for potential coTF interactors, which is also consistent with the 49% overlap between the 419 GO coTF with either the NURSA coTFs or the PCS complex subunits (Fig. 3).

The most stringent case, where proteins must be present in both the manually curated CORUM database and the machine-built hu.MAP database, yielded 667, 1096 and 1183 putative coTFs, for the three bait sets, respectively (Fig. 2H). Of those, 480, 717 and 795 were also found as interaction partners of dbTFs in the BioGRID and IntAct databases (comparing Fig. 2H and F). For comparison, the three bait sets themselves include 285, 576 and 1275 proteins that interact directly with a dbTF in BioGRID and/or IntAct, including 704 dbTFs that interact in a binary fashion with a second dbTF.

3.2. Inclusion of putative coTFs by mining the BioGRID and IntAct databases

When successively applying our three increasingly larger bait sets, up to 1183 proteins were recovered in both hu.MAP and CORUM (Fig. 2H) and are thus considered for manual annotation to transcription coregulator activity GO terms by virtue of residing in a complex that harbours known or presumed coTFs. On the other hand, 1678, 2322 and 2749 proteins were only detected in one of the two protein complex resources (Fig. 2G). Since BioGRID and IntAct contain a total of 5205 human proteins that interact with a known dbTF according to the selection criteria outlined in the Material and Methods (see Section 2.4), we elected to 'rescue' the proteins present in only the CORUM or the hu.MAP database if they also interact with any one of the 1457 known dbTFs according to BioGRID or IntAct (compare Fig. 2G and F). This yielded an extra 818, 1000 and 1248 proteins, respectively (Fig. 2I). We propose to add these to the putative coTF list for future dedicated annotation to transcription coactivator or corepressor molecular activity GO terms.

3.3. Blacklisting of spliceosome, ribosome and proteasome subunits

We identified 1927 'prey' proteins that are neither annotated as dbTF nor as coTF in the Gene Ontology. We denote these putative coTFs. However, they may not all deserve annotation as a coTF. One reason for this may be due to the three largest protein complexes present in the nucleus, namely, the spliceosome, the pre-ribosome and the proteasome. These complexes are not only large, they are also very abundant, and therefore their subunits routinely 'contaminate' mass-spectrometry experiments conducted on nuclear protein extracts. Furthermore, because, by our matrix expansion strategy, a single subunit can bait all the subunits of a protein complex, these large nuclear complexes are easily retrieved. Discarding bait proteins that are known parts of these complexes might be one way of avoiding mining these complexes. However, this would arbitrarily discard moonlighting proteins, involved in both transcription regulation and splicing, translation or protein degradation, of which examples are known [37–40]. Instead, we applied 'post-hoc' blacklist filters based on 15 spliceosome, ribosome or proteasome *cellular component* terms and the *biological process* term 'mRNA splicing, via spliceosome' (see Section 2.8). This blacklisted 293 proteins recovered with bait set 1 and 407 when bait set 2 was used. None of the additional 166 putative coTF proteins recovered when bait set 3 was expanded with 1457 dbTFs were blacklisted (SupTab1).

3.4. List of putative coTFs for targeted annotation

Of the final list of 3803 proteins resulting from the screening of complexes and molecular interactions for putative coTFs, 893 are known or presumed coTFs that we used as baits and 1457 are dbTFs (Fig. 2A). As there were seven known dbTFs in the NURSA bait set and two in the remodelling enzymes set, this leaves 1462 (3803-893-1457+9) putative coTF proteins. Of these, 369 are on the blacklist, resulting in a grand total of 1093 putative coTFs that are neither splicing factors nor ribosome or proteasome subunits. Of these 1093 proteins, 868 reside in at least one selected putative coTF protein complex in hu.MAP and 558 in a CORUM complex, indicating that both complex databases contributed uniquely to our mining output. Notably, only 139 of these 1093 putative coTFs do not interact with one of the 1457 dbTFs in the BioGRID and IntAct databases. Our approach has thus identified 1093 as yet unannotated - putative - human coTFs protein complex subunits (SupTab1).

3.5. The coverage of putative coTFs in the Complex Portal, CORUM and hu.MAP databases

As discussed above, some of the putative coTFs may be false positive hits recovered from large-scale experiments rather than validated functional screens. We therefore assessed their likelihood of being true coTFs by checking their membership in a third party of well annotated complexes, namely those harboured by the Complex Portal at EBI. The Complex Portal is emerging as the authoritative database for macromolecular complexes. It is a manually curated, encyclopaedic resource of macromolecular complexes from a range of model organisms [28].

In order to assess both the comprehensiveness and the quality of the Complex Portal, for every protein on the final list resulting from bait set 3, it was determined whether it is also recovered from the Complex Portal by mining it with the same bait proteins. For 587 proteins this is the case. The majority of these had been mined from both hu.MAP and CORUM (Fig. 4A). Additionally, when for every unique combination of 'bait protein – prey protein' that occurs on the final list we check in which sources this combination was found, the combinations found in the Complex Portal largely overlap with those found in hu.MAP and CORUM (Fig. 4B). Altogether these results show that the Complex Portal currently includes fewer complexes than other sources, but that those complexes are reliable, as they are included in both CORUM and hu.MAP. Indeed, 98% of the putative coTFs found in Complex Portal are also in CORUM and hu.MAP, while 90% of our bait-prey interactions that were found in the Complex Portal are also found in at least one other complex resource with most found in both. On the other hand, 1348 putative coTFs are not found in Complex Portal (Fig. 4A). However, the fact that the overlap of bait-prey interaction pairs is smaller than that for

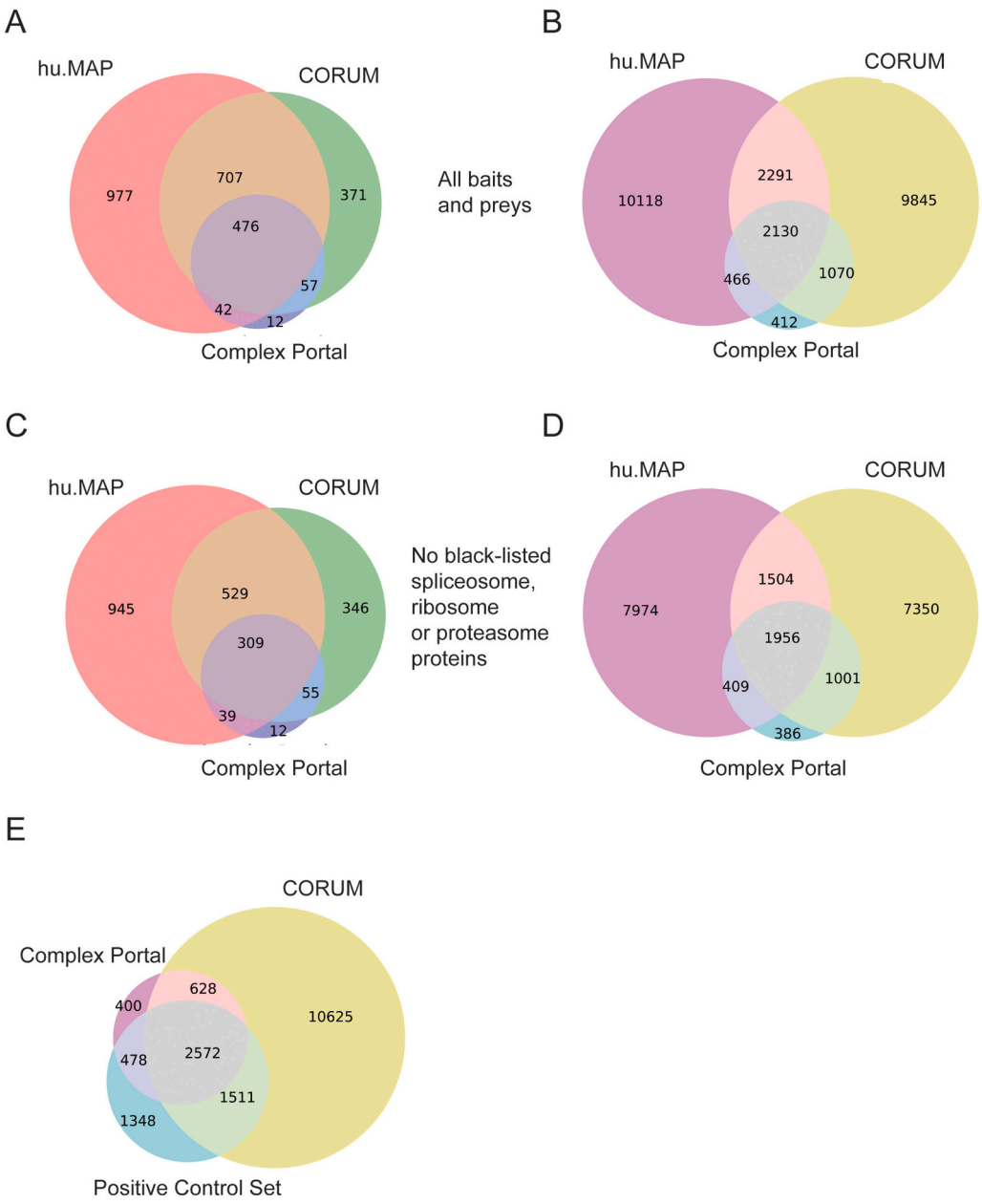


Fig. 4. A. Number of proteins and B. number of bait-prey combinations obtained with bait set 3 that are found in hu.MAP, CORUM and the Complex Portal. C. and D. As A. and B., but upon omission of blacklisted spliceosomal, ribosomal and proteasomal proteins. E. Overlap of the number of all possible intra-complex coTF interactions in the manually curated CORUM and Complex Portal databases and the 'in-house' positive control complex set (PCS).

proteins (Fig. 4B, D), suggests that in some cases the shared proteins are found in different complexes in the three resources.

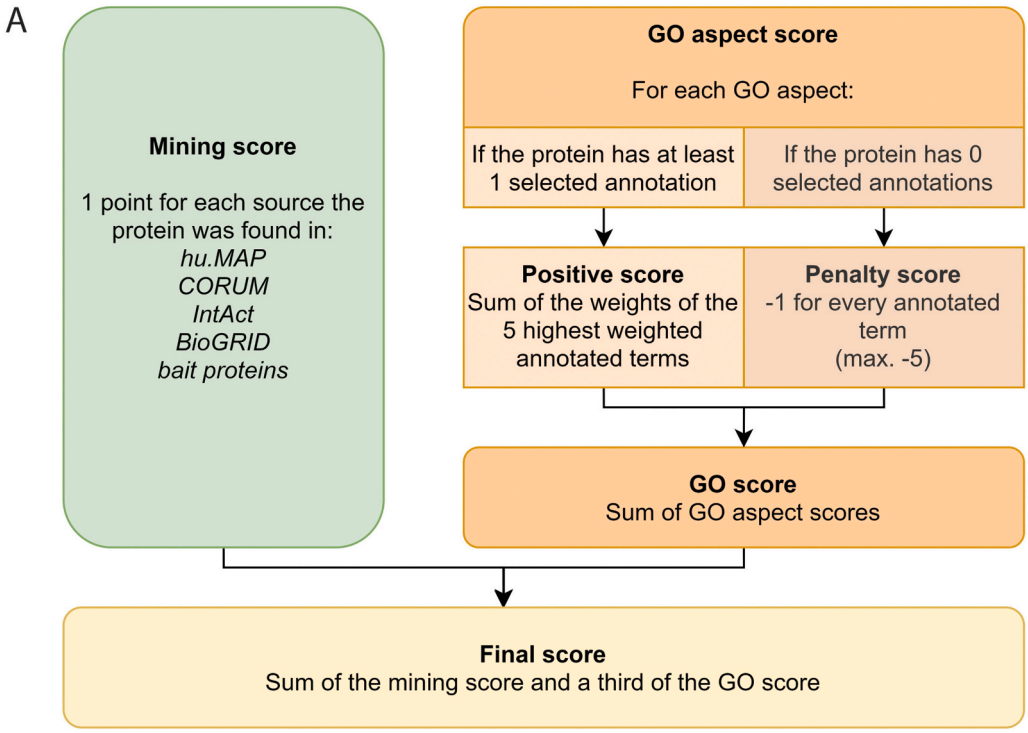
Removal of the 407 blacklisted proteins (Fig. 4D, see Section 3.3) from the input reduced the number of protein-protein interactions commensurably (Fig. 4E). The number of interactions not covered by the Complex Portal, but covered by both hu.MAP and CORUM, was reduced from 2291 to 1504, leaving 529 putative coTFs (Fig. 4C) as being of particular interest for further curation by virtue of not yet being represented in the Complex Portal. Moreover, the Complex Portal complexes already harbour 415 putative coTFs. This provides two distinct starting points for coTF curation reviews.

Finally, and to our satisfaction, the Complex Portal already harbours 51% of our in-house positive control complex set interactions and for CORUM this is 69% (Fig. 4E). This provides a third in-road for an annotation review of putative coTF proteins that are present in canonical transcription coTF complexes.

3.6. *Devising prioritisation scoring systems*

The list of 3803 known and putative human transcription regulators might contain a substantial number of false positive hits because all sources, except CORUM, contain experimental evidence from high-throughput experiments that potentially include artefactual interactions. To enable prioritisation of proteins for manual curation, a scoring system was developed. This score is built-up from a mining score and a GO score. An overview of the scoring strategy is shown in Fig. 5.

The mining score is a score between 0 and 5 that counts in how many of the starting data sources (bait proteins, hu.MAP, CORUM, IntAct and BioGRID) each potential transcription regulator protein was found (Fig. 5, see Section 2.6). The GO score is the sum of a positive score that reflects existing GO aspect (BP, MF or CC) annotations to terms consistent with a role in transcription regulation and, failing this, of GO aspect penalty scores for proteins that are well-annotated with GO terms but



B

$$weight = \frac{n_{annotated\ coregulators}}{n_{total\ coregulators} \times 2} + 0.5$$

Fig. 5. A. Overview of the scoring system applied to score the likeliness of each protein being a coTF. B. Formula used to assign a weight to each selected GO term based on the fraction of known GO coTF proteins associated with that term; *nannotated coregulators* is the number of coregulators annotated with a specific GO term; *ntotal coregulators* represents the total number of proteins annotated as having 'transcription coregulator activity' (GO:0003712 and descendant terms).

not to terms that indicate a potential role in gene transcription. In this system, proteins that do not have any, or hardly any annotations, will get neither a high nor a low GO score, thereby circumventing biases caused by intense research and annotation efforts of certain proteins. To obtain a global score, the GO score was first divided by 3 to make it fall within the range −5 to +5, which is comparable to the mining score's range (0 to 5). Adding the GO and mining scores provides a final score with the theoretical range of −5 to 10 (Fig. 5, see Section 2.7).

To get an impression of the distribution of the GO scores, mining scores, and final scores, the scores were computed for a number of subsets (Fig. 6). Proteins on the bait lists generally scored higher than newly found, and putative coTFs. The very broad range of the scores, however, shows that in all sets there is a substantial number of preys with high scores that may represent 'low-hanging fruits' for annotation as coTFs (Fig. 6A, SupTab1).

In order to see how well the GO scores and mining scores agree, the distribution of GO scores was plotted separately for each possible mining score (Fig. 6B). This shows that higher mining scores generally correspond with higher GO scores. However, this does not hold true for mining scores of 0 and 1 because proteins with a mining score of 0 are known dbTFs and those with a mining score of 1 are proteins that made the list only for being a bait protein. Hence, these are proteins that are already known to have some kind of function in transcription regulation and will therefore have a higher GO score. To examine this correlation further, it was repeated for different subsets of the list. Fig. 6C shows that the GO scores and the mining scores correlate well for putative coTFs. As expected, the GO scores of known transcription factors and their coregulators are higher and only increase marginally as a function of the mining score. We therefore suggest that interested researchers focus on putative coTFs with either high mining scores, for which protein complexes are established in the literature, or high GO aspect scores, for which much literature-based functional annotation has already been performed (Fig. 6C). This indicates that the usefulness of one or the other partial score for coTF annotation is most powerful when considering also the other partial score.

Finally, we analysed the respective contributions of the five data sources (baits, CORUM, hu.MAP, BioGRID, IntAct). The GO scores were plotted for mining scores of 3 and 4, by grouping them by data source (Fig. 7). Proteins that are not on one of the bait lists generally have lower GO scores. This is not necessarily because they are not involved in transcription regulation, but may be because they are not yet documented as such. Proteins that are not found in the machine-generated hu.MAP complexes generally have higher GO scores which may be related to the propensity of a higher proportion of false positive preys in this data set. Proteins that are not baits and that are not found in hu.MAP complexes show the widest distribution of scores. As BioGRID, IntAct and CORUM are all manually and independently curated this demonstrates the power and value of manual curation.

4. Discussion

It has been estimated that 30 to 50% of all the proteins in a simple eukaryotic cell such as the unicellular budding yeast model organism form stable complexes [41]. Furthermore, some of the complexes involved in eukaryotic transcription are assembled by specific chaperonin subunits such as Bud27 and its human paralog URI1 [42–44]. The ability to copurify through multiple chromatographic steps is an imperfect but rather extended biochemical standard to call a set of proteins the subunits of any particular protein complex [41,45,46], even though these associations are often incomplete, spurious or might

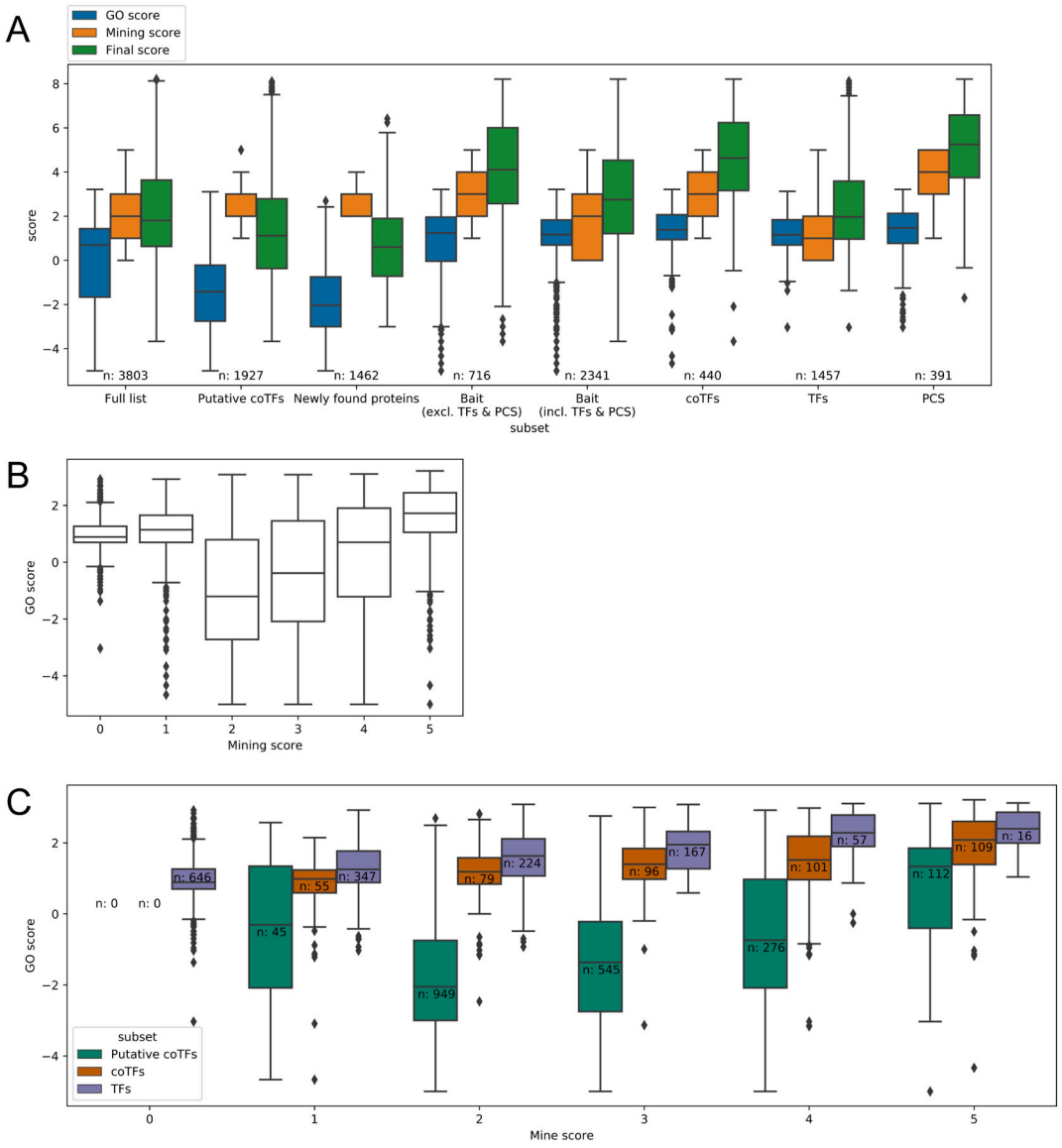


Fig. 6. A. Distributions of final scores, GO scores and mining scores given subsets of proteins on the final list made using bait set 3. B. Distributions of GO scores given to proteins on the final list made using bait set 3 for each possible mining score. C. Distributions of GO scores given to subsets of proteins on the final list made using bait set 3 for each possible mining score.

represent a mix of different complexes. A second standard to define the subunits of a protein complex involves genetic epistasis of their loss of function alleles, which strongly, but not perfectly, correlates with complex participant co-occurrence [47,48]. Furthermore, to date, it remains unclear to what extent the rather large protein complexes involved in modulating transcription are themselves remodelled as they perform their function in vivo, even though post-translational modification patterns probably indicate meta-labile conformations. Nuclear proteins and protein complexes transiently bump into each other and associate with DNA in the nucleoplasm, possibly in phase-separated liquid nanodroplets [49–52]. When such binding events are specific and last long enough, this may have functional consequences, such as the modulation of gene transcription.

The difference between transient protein-protein interactions and stable interactions as the basis of complexes played an important role in our approach. Short-lived interactions may direct complex assembly, complex chaperoning or recruitment and activation of coTF complexes by dbTFs. However, there are also examples in the literature, where dbTFs are integral parts of stable coTF complexes [53–56]. In this light, the large number of interactions found in CORUM but not in hu.MAP could indicate the manual curation of transient interactions that are derived from multiple types of evidence which are not reflected by the programmatically-inferred hu.MAP set of protein complexes which are only based on affinity purification and cofractionation experiments. The nature of interactions is also relevant in the context of the observation that the dbTFs baits yielded relatively few extra proteins in CORUM and/or hu.MAP when compared to bait set 2, which consists of known and strongly presumed coTFs. Considering this, it might be interesting to take a closer look at the remaining dbTF interactors in IntAct and BioGRID, which could add transient dbTF interactors to the list. However, as transcription factors are often endpoints of signaling pathways, that approach might also yield 'upstream regulators' of dbTFs rather than the coTFs that act in conjunction with the DNA-bound dbTFs at enhancers and/or gene promoters in order to influence gene-specific transcription levels.

In this study we compiled lists of human proteins known or suspected to be involved in transcription regulation by virtue of their physical association with a known transcription regulator. The final list includes

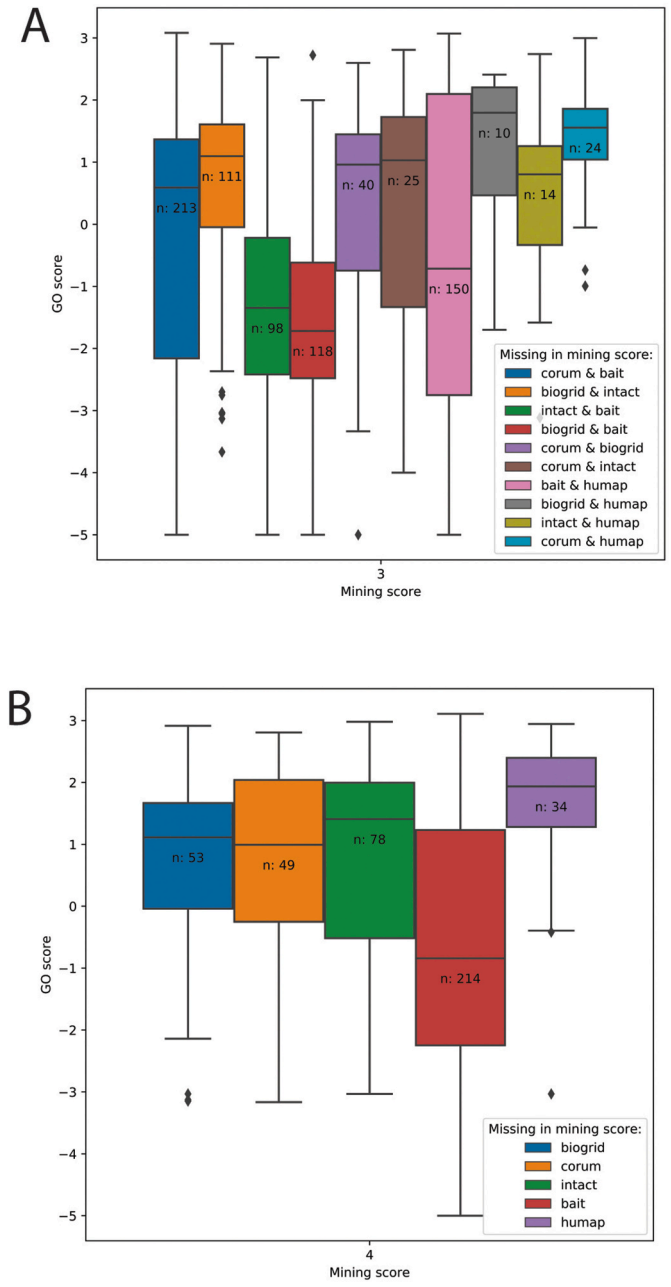


Fig. 7. Distributions of GO scores awarded to proteins on the final list using bait set 3. The proteins are grouped on the data sources the proteins were not found in. A. Analysis of proteins with a mining score of 3. B. Analysis of proteins with a mining score of 4.

1927 proteins that are presently not annotated as dbTF or coTF. Of the 1462 newly found proteins that are not in our three bait sets, 369 are annotated to spliceosome, ribosome or proteasome and these were therefore blacklisted. This leaves 1093 putative coTF proteins to consider for manual curation alongside the 474 strongly presumed coTFs from bait-set 2 (Fig. 2A). Four proteins with high total scores, CTNNB1, HDAC2, KAT8 and WDR5, highlight the suitability of the scoring system for selecting putative coTFs for further manual annotation purposes: During the time between analysis and revision of this manuscript, CTNNB1 and KAT8 were annotated in GO as coTFs. Furthermore, HDAC2 and WDR5 are involved in chromatin modification, the former directly through its enzymatic activity, the latter by association with chromatin modifying complexes [17,18]. The scoring system we present can thus be used to further prioritize proteins for manual annotation. An even larger list of putative coTFs could be created if we relaxed the rule that an interactor of a dbTF in BioGRID or IntAct must be present in a hu.MAP or CORUM protein complex. Even though this would greatly expand the potential coTF set (Fig. 2F), the scoring system we present might be able to deal with those. All the proteins that interact with a dbTF could then be thoroughly analysed.

Altogether, we conclude that GO does not yet comprehensively cover the molecular function annotation of human coTFs. Our lists and prioritisation scores should therefore be useful for upcoming coTF annotation efforts. Since the putative coTFs we identified reside in protein complexes, we propose to concomitantly annotate coTF protein complexes in the Complex portal [28]. Moreover, we call on the research community members who devise computational and experimental approaches that provide data on known, presumed and putative coTFs, to approach GO through the help desk to indicate relevant publications and data sets that could be used for functional curation [Cortés et al., BBA this issue] of dbTFs, coTFs and the complexes they reside in.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbagrm.2021.194749.

Funding

This work was supported by the GREEKC COST action CA1502. We acknowledge funding from the National Eye Institute, National Human Genome Research Institute, National Heart, Lung, and Blood Institute, National Institute of Allergy and Infectious Diseases, National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of General Medical Sciences, National Cancer Institute, National Institute On Aging, and National Institute of Mental Health of the National Institutes of Health [U24HG007822]; National Human Genome Research Institute [U41HG002273]; National Institute of General Medical Sciences [R01GM080646, P20GM103446] (the content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health), Open Targets (grant agreements OTAR-044 and OTAR02-048) and the Wellcome Trust grant INVAR (grant ref: 212925/Z/18/Z) and European Molecular Biology Laboratory core funds.

Credit authorship contribution statement

All authors have read and proofed the manuscript and its associated files.

Declaration of competing interest

All the authors of this manuscript declare that they have no competing interests related to this manuscript.

Acknowledgements

We are grateful to the members of GREEKC for useful discussions and to Dr. Simon van Heeringen and all the members of his research team for their help with coding. Thanks also to Dr. Pascal Vos for his preliminary work and to Ruth Lovering and Pascale Gaudet for their precious input in the course of this work.

References

[1] S. Sainsbury, C. Bernecky, P. Cramer, Structural basis of transcription initiation by RNA polymerase II, Nat. Rev. Mol. Cell Biol. 16 (2015) 129–143.
[2] S. Hahn, Structure and mechanism of the RNA polymerase II transcription machinery, Nat. Struct. Mol. Biol. 11 (2004) 394–403.
[3] K. Gupta, D. Sari-Ak, M. Haffke, S. Trowitzsch, I. Berger, Zooming in on transcription preinitiation, J. Mol. Biol. 428 (2016) 2581–2591.
[4] E.M. Wissink, A. Vihervaara, N.D. Tippens, J.T. Lis, Nascent RNA analyses: tracking transcription and its regulation, Nat. Rev. Genet. 20 (2019) 705–723.

[5] J. Zheng, J.J. Benschop, M. Shales, P. Kemmeren, J. Greenblatt, G. Cagney, F. Holstege, H. Li, N.J. Krogan, Epistatic relationships reveal the functional organization of yeast transcription factors, Mol. Syst. Biol. 6 (2010) 420.
[6] M. Levine, R. Tjian, Transcription regulation and animal diversity, Nature 424 (2003) 147–151.
[7] D.S. Latchman, Transcription factors: an overview, Int. J. Biochem. Cell Biol. 29 (1997) 1305–1312.
[8] H. Szutorisz, N. Dillon, L. Tora, The role of enhancers as centres for general transcription factor recruitment, Trends Biochem. Sci. 30 (2005) 593–599.
[9] D.S. Latchman, Inhibitory transcription factors, Int. J. Biochem. Cell Biol. 28 (1996) 965–974.
[10] F. Spitz, E.E. Furlong, Transcription factors: from enhancer binding to developmental control, Nat. Rev. Genet. 13 (2012) 613–626.
[11] Z.C. Poss, C.C. Ebmeier, D.J. Taatjes, The mediator complex and transcription regulation, Crit. Rev. Biochem. Mol. Biol. 48 (2013) 575–608.
[12] J. Soutourina, Transcription regulation by the mediator complex, Nat. Rev. Mol. Cell Biol. 19 (2018) 262–274.
[13] T.C. Voss, G.L. Hager, Dynamic regulation of transcriptional states by chromatin and transcription factors, Nat. Rev. Genet. 15 (2014) 69–81.
[14] N. Yudkovsky, C. Logie, S. Hahn, C.L. Peterson, Recruitment of the SWI/SNF chromatin remodeling complex by transcriptional activators, Genes Dev. 13 (1999) 2369–2374.
[15] C. Logie, C.L. Peterson, Catalytic activity of the yeast SWI/SNF complex on reconstituted nucleosome arrays, EMBO J. 16 (1997) 6772–6782.
[16] P. Zhang, K. Torres, X. Liu, C.G. Liu, R.E. Pollock, An overview of chromatin-regulating proteins in cells, Curr. Protein Pept. Sci. 17 (2016) 401–410.
[17] M. Lawrence, S. Daujat, R. Schneider, Lateral thinking: how histone modifications regulate gene expression, Trends Genet. 32 (2016) 42–56.
[18] C.R. Clapier, B.R. Cairns, The biology of chromatin remodeling complexes, Annu. Rev. Biochem. 78 (2009) 273–304.
[19] J.J. van Vugt, M. de Jager, M. Murawska, A. Brehm, J. van Noort, C. Logie, Multiple aspects of ATP-dependent nucleosome translocation by RSC and Mi-2 are directed by the underlying DNA sequence, PLoS One 4 (2009), e6345.
[20] S. Venkatesh, J.L. Workman, Histone exchange, chromatin structure and the regulation of transcription, Nat. Rev. Mol. Cell Biol. 16 (2015) 178–189.
[21] D. Schubeler, Function and information content of DNA methylation, Nature 517 (2015) 321–326.
[22] H. Zhu, G. Wang, J. Qian, Transcription factors as readers and effectors of DNA methylation, Nat. Rev. Genet. 17 (2016) 551–565.
[23] C. Gene Ontology, The Gene Ontology resource: enriching a GOld mine, Nucleic Acids Res. 49 (2021) D325–D334.
[24] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, Nat. Genet. 25 (2000) 25–29.
[25] D. Wilson, V. Charoensawan, S.K. Kummerfeld, S.A. Teichmann, DBD—taxonomically broad transcription factor predictions: new content and functionality, Nucleic Acids Res. 36 (2008) D88–D92.
[26] E. Bornberg-Bauer, E. Rivals, M. Vingron, Computational approaches to identify leucine zippers, Nucleic Acids Res. 26 (1998) 2740–2746.
[27] D.N. Messina, J. Glasscock, W. Gish, M. Lovett, An ORFeome-based analysis of human transcription factor genes and the construction of a microarray to interrogate their expression, Genome Res. 14 (2004) 2041–2047.
[28] B.H.M. Meldal, A.J.H. Bye, L. Gajdos, Z. Hammerova, A. Horackova, F. Melicher, L. Perfetto, D. Pokorny, M.R. Lopez, A. Turkova, E.D. Wong, Z. Xie, E.B. Casanova, N. Del-Toro, M. Koch, P. Porras, H. Hermjakob, S. Orchard, Complex Portal 2018: extended content and enhanced visualization tools for macromolecular complexes, Nucleic Acids Res. 47 (2019) D550–D558.
[29] N.J. McKenna, A.J. Cooney, F.J. DeMayo, M. Downes, C.K. Glass, R.B. Lanz, M. A. Lazar, D.J. Mangelsdorf, D.D. Moore, J. Qin, D.L. Steffen, M.J. Tsai, S.Y. Tsai, R. Yu, R.N. Margolis, R.M. Evans, B.W. O'Malley, Minireview: evolution of NURSA, the Nuclear Receptor Signaling Atlas, Mol. Endocrinol. 23 (2009) 740–746.
[30] R.N. Margolis, R.M. Evans, B.W. O'Malley, N.A. Consortium, The Nuclear Receptor Signaling Atlas: development of a functional atlas of nuclear receptors, Mol. Endocrinol. 19 (2005) 2433–2436.
[31] M.J. Koster, B. Snel, H.T. Timmers, Genesis of chromatin and transcription dynamics in the origin of species, Cell 161 (2015) 724–736.
[32] K. Drew, C. Lee, R.L. Huizar, F. Tu, B. Borgeson, C.D. McWhite, Y. Ma, J. B. Wallingford, E.M. Marcotte, Integration of over 9,000 mass spectrometry experiments builds a global map of human protein complexes, Mol. Syst. Biol. 13 (2017) 932.
[33] M. Giurgiu, J. Reinhard, B. Brauner, I. Dunger-Kaltenbach, G. Fobo, G. Frishman, C. Montrone, A. Ruepp, CORUM: the comprehensive resource of mammalian protein complexes-2019, Nucleic Acids Res. 47 (2019) D559–D563.
[34] C. Stark, B.J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, M. Tyers, BioGRID: a general repository for interaction datasets, Nucleic Acids Res. 34 (2006) D535–D539.
[35] S. Orchard, M. Ammari, B. Aranda, L. Breuza, L. Briganti, F. Broackes-Carter, N. H. Campbell, G. Chavali, C. Chen, N. del-Toro, M. Duesbury, M. Dumousseau, E. Galeota, U. Hinz, M. Iannuccelli, S. Jagannathan, R. Jimenez, J. Khadake, A. Lagreid, L. Licata, R.C. Lovering, B. Meldal, A.N. Melidoni, M. Milagros, D. Peluso, L. Perfetto, P. Porras, A. Raghunath, S. Ricard-Blum, B. Roechert, A. Stutz, M. Tognolli, K. van Roey, G. Cesareni, H. Hermjakob, The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases, Nucleic Acids Res. 42 (2014) D358–D363.
[36] J.M. Villaveces, R.C. Jimenez, P. Porras, N. Del-Toro, M. Duesbury, M. Dumousseau, S. Orchard, H. Choi, P. Ping, N.C. Zong, M. Askenazi, B. H. Habermann, H. Hermjakob, Merging and scoring molecular interactions utilising existing community standards: tools, use-cases and a case study, Database (Oxford) 2015 (2015).
[37] X.H. Lei, X. Shen, X.Q. Xu, H.S. Bernstein, Human Cdc5, a regulator of mitotic entry, can act as a site-specific DNA binding protein, J. Cell Sci. 113 (Pt 24) (2000) 4523–4531.
[38] M. Grote, E. Wolf, C.L. Will, I. Lemm, D.E. Agafonov, A. Schomburg, W. Fischle, H. Urlaub, R. Luhrmann, Molecular architecture of the human Prp19/CDC5L complex, Mol. Cell. Biol. 30 (2010) 2105–2119.
[39] E. Ezhkova, W.P. Tansey, Proteasomal ATPases link ubiquitylation of histone H2B to methylation of histone H3, Mol. Cell 13 (2004) 435–442.
[40] A. Ferdous, F. Gonzalez, L. Sun, T. Kodadek, S.A. Johnston, The 19S regulatory particle of the proteasome is required for efficient transcription elongation by RNA polymerase II, Mol. Cell 7 (2001) 981–991.
[41] S.R. Collins, P. Kemmeren, X.C. Zhao, J.F. Greenblatt, F. Spencer, F.C. Holstege, J. S. Weissman, N.J. Krogan, Toward a comprehensive atlas of the physical interactome of Saccharomyces cerevisiae, Mol. Cell. Proteomics 6 (2007) 439–450.
[42] M.C. Miron-Garcia, A.I. Garrido-Godino, V. Garcia-Molinero, F. Hernandez-Torres, S. Rodriguez-Navarro, F. Navarro, The prefoldin bud27 mediates the assembly of the eukaryotic RNA polymerases in an rpb5-dependent manner, PLoS Genet. 9 (2013), e1003297.
[43] G. Millan-Zambrano, A. Rodriguez-Gil, X. Penate, L. de Miguel-Jimenez, M. Morillo-Huesca, N. Krogan, S. Chavez, The prefoldin complex regulates chromatin dynamics during transcription elongation, PLoS Genet. 9 (2013), e1003776.
[44] M.C. Miron-Garcia, A.I. Garrido-Godino, V. Martinez-Fernandez, A. Fernandez-Pevida, A. Cuevas-Bermudez, M. Martin-Exposito, S. Chavez, J. de la Cruz, F. Navarro, The yeast prefoldin-like URI-orthologue Bud27 associates with the RSC nucleosome remodeler and modulates transcription, Nucleic Acids Res. 42 (2014) 9666–9676.
[45] R.G. Roeder, W.J. Rutter, Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms, Nature 224 (1969) 234–237.
[46] J.D. Dignam, P.L. Martin, B.S. Shastry, R.G. Roeder, Eukaryotic gene transcription with purified components, Methods Enzymol. 101 (1983) 582–598.
[47] S.R. Collins, K.M. Miller, N.L. Maas, A. Roguev, J. Fillingham, C.S. Chu, M. Schuldiner, M. Gebbia, J. Recht, M. Shales, H. Ding, H. Xu, J. Han, K. Ingvarsdottir, B. Cheng, B. Andrews, C. Boone, S.L. Berger, P. Hieter, Z. Zhang, G.W. Brown, C.J. Ingles, A. Emili, C.D. Allis, D.P. Toczyski, J.S. Weissman, J. F. Greenblatt, N.J. Krogan, Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map, Nature 446 (2007) 806–810.
[48] M. Costanzo, B. VanderSluis, E.N. Koch, A. Baryshnikova, C. Pons, G. Tan, W. Wang, M. Usaj, J. Hanchard, S.D. Lee, V. Pelechano, E.B. Styles, M. Billmann, J. van Leeuwen, N. van Dyk, Z.Y. Lin, E. Kuzmin, J. Nelson, J.S. Piotrowski, T. Srikumar, S. Bahr, Y. Chen, R. Deshpande, C.F. Kurat, S.C. Li, Z. Li, M.M. Usaj, H. Okada, N. Pascoe, B.J. San Luis, S. Sharifpoor, E. Shuteriqi, S.W. Simpkins, J. Snider, H.G. Suresh, Y. Tan, H. Zhu, N. Malod-Dognin, V. Janjic, N. Przulj, O. G. Troyanskaya, I. Stagljar, T. Xia, Y. Ohya, A.C. Gingras, B. Raught, M. Boutros, L. M. Steinmetz, C.L. Moore, A.P. Rosebrock, A.A. Caudy, C.L. Myers, B. Andrews, C. Boone, A global genetic interaction network maps a wiring diagram of cellular function, Science 353 (2016).
[49] I. Kwon, M. Kato, S. Xiang, L. Wu, P. Theodoropoulos, H. Mirzaei, T. Han, S. Xie, J. L. Corden, S.L. McKnight, Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains, Cell 155 (2013) 1049–1060.
[50] K.A. Burke, A.M. Janke, C.L. Rhine, N.L. Fawzi, Residue-by-residue view of in vitro FUS granules that bind the C-terminal domain of RNA polymerase II, Mol. Cell 60 (2015) 231–241.
[51] B.R. Sabari, A. Dall'Agnese, A. Boija, I.A. Klein, E.L. Coffey, K. Shrinivas, B. J. Abraham, N.M. Hannett, A.V. Zamudio, J.C. Manteiga, C.H. Li, Y.E. Guo, D. S. Day, J. Schuijers, E. Vasile, S. Malik, D. Hnisz, T.I. Lee, I.I. Cisse, R.G. Roeder, P. A. Sharp, A.K. Chakraborty, R.A. Young, Coactivator condensation at super-enhancers links phase separation and gene control, Science 361 (2018).
[52] A.V. Zamudio, A. Dall'Agnese, J.E. Henninger, J.C. Manteiga, L.K. Afeyan, N. M. Hannett, E.L. Coffey, C.H. Li, O. Oksuz, B.R. Sabari, A. Boija, I.A. Klein, S. W. Hawken, J.H. Spille, T.M. Decker, I.I. Cisse, B.J. Abraham, T.I. Lee, D.J. Taatjes, J. Schuijers, R.A. Young, Mediator condensates localize signaling factors to key cell identity genes, Mol. Cell 76 (2019) 753–766 e756.
[53] B.R. Cairns, A. Schlichter, H. Erdjument-Bromage, P. Tempst, R.D. Kornberg, F. Winston, Two functionally distinct forms of the RSC nucleosome-remodeling complex, containing essential AT hook, BAH, and bromodomains, Mol. Cell 4 (1999) 715–723.
[54] M.L. Angus-Hill, A. Schlichter, D. Roberts, H. Erdjument-Bromage, P. Tempst, B. R. Cairns, A Rsc3/Rsc30 zinc cluster dimer reveals novel roles for the chromatin remodeler RSC in gene expression and cell cycle control, Mol. Cell 7 (2001) 741–751.
[55] C. Campsteijn, A.M. Wijnands-Collin, C. Logie, Reverse genetic analysis of the yeast RSC chromatin remodeler reveals a role for RSC3 and SNF5 homolog 1 in ploidy maintenance, PLoS Genet. 3 (2007), e92.
[56] G. Badis, E.T. Chan, H. van Bakel, L. Pena-Castillo, D. Tillo, K. Tsui, C.D. Carlson, A. J. Gossett, M.J. Hasinoff, C.L. Warren, M. Gebbia, S. Talukder, A. Yang, S. Mnaimneh, D. Terterov, D. Coburn, A. Li Yeo, Z.X. Yeo, N.D. Clarke, J.D. Lieb, A. Z. Ansari, C. Nislow, T.R. Hughes, A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters, Mol. Cell 32 (2008) 878–887.